

Compte rendu TD1 : XML et DTDs

BRESSAND - ABADIE

1) Des schémas aux données

DTD 1 :

Version 1 :

```
<?xml version="1.0" ?>
<presse>
  <journal>
    <article titre="Science et vie" auteur="001">
      <corps>
        Numéro spécial sur les technologies du web
      </corps>
    </article>
    <article titre="Nature" auteur="002">
      <corps>
        Numéro spécial sur le big data
      </corps>
    </article>
  </journal>
  <journalistes>
    <journaliste id="001">
      <anonyme />
    </journaliste>
    <journaliste id="002">
      <pseudonyme>Spirou</pseudonyme>
    </journaliste>
  </journalistes>
</presse>
```

Version 2 :

```
<?xml version="1.0"?>
```

```
<presse>
```

```
<journal>
```

```
<article titre="DTD1" auteur="auth1">
```

```
<corps>
```

Article respectant la DTD1 du TP1 de représentation
des données du web.

```
</corps>
```

```
</article>
```

```
</journal>
```

```
<journalistes>
```

```
<journaliste idJ="auth1">
```

```
<anonyme></anonyme>
```

```
</journaliste>
```

```
<journaliste idJ="auth2">
```

```
<nom>Obama</nom>
```

```
<prenom>Barack</prenom>
```

```
</journaliste>
```

```
</journalistes>
```

```
</presse>
```

DTD 2 :

Version 1 :

```
<?xml version="1.0" ?>
```

```
<batiment>
```

```
<etage>
```

```
<description>
```

```
Etage recherche et developpement
```

```
</description>
```

```
<bureau>
```

```
<code>Banana</code>
```

```
<personne>Jean-jacques</personne>
```

```
<personne>Thérèse</personne>
```

```
</bureau>
```

```
<bureau>
```

```
<code>Potatoe</code>
```

```
</bureau>
```

```
</etage>
```

```
</batiment>
```

Version 2 :

```
<?xml version="1.0"?>
<batiment>
  <etage>
    <description>
      Rez-de-chaussez
    </description>
    <salle>
      <nombrePlaces>40</nombrePlaces>
    </salle>
    <salle>
      <nombrePlaces>40</nombrePlaces>
    </salle>
    <salle>
      <nombrePlaces>40</nombrePlaces>
    </salle>
  </etage>
  <etage>
    <description>
      Premier étage
    </description>
    <salle>
      <nombrePlaces>50</nombrePlaces>
    </salle>
    <salle>
      <nombrePlaces>30</nombrePlaces>
    </salle>
    <salle>
      <nombrePlaces>25</nombrePlaces>
    </salle>
  </etage>
</batiment>
```

2) Des données aux schémas

XML 1 :

Version 1 :

```
<!DOCTYPE C[
  <!ELEMENT C (B, B, A, C, B, B, B, A)? >
  <!ELEMENT B EMPTY>
  <!ELEMENT A EMPTY>
]>
```

Version 2 :

```
<!DOCTYPE C [  
  <!ELEMENT C (A|B|C)*>  
  <!ELEMENT A EMPTY>  
  <!ELEMENT B EMPTY>  

```

XML 2 :

Version 1 :

```
<!DOCTYPE C[  
  <!ELEMENT C ( B, C+, A, D+, A )? >  
  <!ELEMENT B EMPTY>  
  <!ELEMENT A EMPTY>  
  <!ELEMENT D EMPTY>  

```

Version 2 :

```
<!DOCTYPE C [  
  <!ELEMENT C ( (B*, C*, A, D*, A)+ )? >  
  <!ELEMENT A EMPTY>  
  <!ELEMENT B EMPTY>  
  <!ELEMENT D EMPTY>  

```

XML 3 :

Version 1 :

```
<!DOCTYPE C[  
  <!ELEMENT C ( B+, C+, B )? >  
  <!ELEMENT B EMPTY>  

```

Version 2 :

```
<!DOCTYPE C[  
  <!ELEMENT C (B|C)* >  
  <!ELEMENT B EMPTY>  

```

XML 4 :

Version 1 :

```
<!DOCTYPE C[
  <!ELEMENT C ( A | B | C | D )* >
  <!ELEMENT A EMPTY>
  <!ELEMENT B ( A | D ) >
  <!ELEMENT D EMPTY>
]>
```

Version 2 :

```
<!DOCTYPE C [
  <!ELEMENT C ((B|C)*|(A|D))>
  <!ELEMENT B (A|D)>
  <!ELEMENT A EMPTY>
  <!ELEMENT D EMPTY>
]>
```

XML 5 :

Version 1 :

```
<!DOCTYPE C [
  <!ELEMENT C ( A | B | C )*>
  <!ELEMENT A EMPTY>
  <!ELEMENT B EMPTY>
]>
```

Version 2 :

```
<!DOCTYPE C [
  <!ELEMENT C EMPTY>
]>
```

XML 6 :

Version 1 :

```
<!DOCTYPE C [
  <!ELEMENT C ( B | C)+>
  <!ELEMENT B #PCDATA>
  <!ATTLIST B id ID #REQUIRED>
  <!ATTLIST B friend CDATA #REQUIRED>
]>
```

Version 2 :

```
<!DOCTYPE C [
  <!ELEMENT C B>
  <!ELEMENT B #PCDATA>
  <!ATTLIST B id ID #REQUIRED>
  <!ATTLIST B friend CDATA #IMPLIED>
]>
```

3) Modélisation : à quoi ressemble un Tweet ?

```
<!DOCTYPE tweeter [  
<!ELEMENT tweeter (tweet+, utilisateur+)>  
  
    <!ELEMENT tweet (post, date, coordonnees?, cptRetweet)>  
    <!ATTLIST tweet idTweet ID #REQUIRED>  
    <!ATTLIST tweet idAuthor IDREF #REQUIRED>  
    <!ATTLIST tweet reponse IDREF #IMPLIED>      <!-- référence vers le tweet précédent-->  
    <!ATTLIST tweet allowAnswers category (true|false) #REQUIRED>  
  
    <!ELEMENT post (message|video|image)+>  
        <!ELEMENT message ((text|hashtag|refUser)+, size, font, color, language)>  
            <!ELEMENT text (#PCDATA)>  
            <!ELEMENT hashtag (#PCDATA)>  
            <!ELEMENT refUser IDREF>  
            <!ELEMENT size (#PCDATA)>  
            <!ELEMENT font (#PCDATA)>  
            <!ELEMENT color (#PCDATA)>  
            <!ELEMENT language (#PCDATA)>  
  
            <!ELEMENT video url>  
            <!ELEMENT image url>  
                <!ELEMENT url (#PCDATA)>  
  
    <!ELEMENT date (secondes, fuseauHoraire)>  
        <!ELEMENT secondes (#PCDATA)>  
        <!ELEMENT fuseauHoraire (#PCDATA)>  
  
    <!ELEMENT coordonnees (gps, ville, pays, os)>  
        <!ELEMENT gps (longitude, latitude)>  
            <!ELEMENT longitude (#PCDATA)>  
            <!ELEMENT latitude (#PCDATA)>  
        <!ELEMENT ville (#PCDATA)>  
        <!ELEMENT pays (#PCDATA)>  
        <!ELEMENT os (#PCDATA)>  
  
    <!ELEMENT cptRetweet (#PCDATA)> <!-- Nb de fois que le tweet est retweeté-->  
  
    <!ELEMENT utilisateur (nom, description, photo, followers)>  
    <!ATTLIST utilisateur idAut ID #REQUIRED>  
        <!ELEMENT nom (#PCDATA)>  
        <!ELEMENT description (#PCDATA)>  
        <!ELEMENT photo (#PCDATA)> <!-- url de l'image -->  
        <!ELEMENT suiveurs (#PCDATA)> <!-- Nb d'abonnés -->  
        <!ELEMENT suivis (#PCDATA)>      <!-- Nb d'utilisateurs auquel cet utilisateur est  
abonné -->  
>
```

4) Synthèse des modèles

```
<!DOCTYPE tweeter [  
<!ELEMENT tweeter (utilisateur* | ( utilisateur+, tweet*)) >  
  
    <!ELEMENT tweet (post, date, coordonnees?, os?, cptRetweet)>  
    <!ATTLIST tweet idTweet ID #REQUIRED>  
    <!ATTLIST tweet idAuthor IDREF #REQUIRED>  
    <!ATTLIST tweet reponse IDREF #IMPLIED>      <!-- référence vers le tweet précédent-->  
    <!ATTLIST tweet allowAnswers category (true|false) #REQUIRED>  
  
    <!ELEMENT post (message|video|image)+>  
        <!ELEMENT message ((text|hashtag|refUser)+, size, font, color, language)>  
            <!ELEMENT text (#PCDATA)>  
            <!ELEMENT hashtag (#PCDATA)>  
            <!ELEMENT refUser IDREF>  
            <!ELEMENT size (#PCDATA)>  
            <!ELEMENT font (#PCDATA)>  
            <!ELEMENT color (#PCDATA)>  
            <!ELEMENT language (#PCDATA)>  
  
            <!ELEMENT video url>  
            <!ELEMENT image url>  
                <!ELEMENT url (#PCDATA)>  
  
    <!ELEMENT date (secondes, fuseauHoraire)>  
        <!ELEMENT secondes (#PCDATA)>  
        <!ELEMENT fuseauHoraire (#PCDATA)>  
  
    <!ELEMENT coordonnees (gps, ville, pays)>  
        <!ELEMENT gps (longitude, latitude)>  
            <!ELEMENT longitude (#PCDATA)>  
            <!ELEMENT latitude (#PCDATA)>  
        <!ELEMENT ville (#PCDATA)>  
        <!ELEMENT pays (#PCDATA)>  
  
    <!ELEMENT os (#PCDATA)>  
    <!ELEMENT cptRetweet (#PCDATA)> <!-- Nb de fois que le tweet est retweeté-->  
  
    <!ELEMENT utilisateur (nom, description, photo, followers)>  
    <!ATTLIST utilisateur idAut ID #REQUIRED>  
        <!ELEMENT nom (#PCDATA)>  
        <!ELEMENT description (#PCDATA)>  
        <!ELEMENT photo (#PCDATA)> <!-- url de l'image -->  
        <!ELEMENT suiveurs (#PCDATA)> <!-- Nb d'abonnés -->  
        <!ELEMENT suivis (#PCDATA)>      <!-- Nb d'utilisateurs auquel cet utilisateur est  
abonné -->  
>  
>
```

5) Cas particuliers des DTDs

DTD1 :

Impossible car l'élément C est défini deux fois

DTD2 :

```
<?xml version="1.0"?>
```

```
<EMPTY/>
```

DTD3 :

Impossible car boucle infinie.

DTD4 :

Impossible car la DTD n'est pas valide, utiliser des "<>" à la place des "()"

DTD5 :

Impossible, on ne peut pas utiliser "EMPTY" dans une expression régulière, utiliser le caractère '?'

DTD6 :

Les éléments C créer d'autres éléments C sans fin car l'étoile ne porte que sur l'élément B